SOME DATA ANALYTIC PROBLEMS IN MODELS WITH LATENT VARIABLES

by

Ulf Olsson

DEPARTMENT OF STATISTICS
UNIVERSITY OF UPPSALA
1978

SOME DATA ANALYTIC PROBLEMS IN MODELS WITH LATENT VARIABLES

ULF OLSSON

Uppsala 1978

ISBN 91-506-0156-3

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41548924_0014

University of Uppsala
Department of Statistics
P.o. Box 513
S-751 20 Uppsala Sweden

Research Report 78 - 8

SOME DATA ANALYTIC PROBLEMS IN MODELS WITH LATENT VARIABLES

Ulf Olsson

Research reported in this paper has been supported in part by the Bank of Sweden Tercentenary Foundation, under project "Structural Equation Models in the Social Sciences", project director Karl G. Jöreskog, and in part by the Swedish Council for Social Science Research under project "Statistical Methods for Analysis of Longitudinal Data", project director Karl G. Jöreskog.

The present paper is a summary of the following communications:

I. Olsson, U. and Bergman, L. R. A longitudinal factor model for studying change in ability structure. Multivariate behavioral research, 1977, 12, 221-242

II. Olsson, U. Selection bias in confirmatory factor analysis. Research report 78-4 , Department of Statistics, University of Uppsala, 1978

III. Olsson, U. A comparison of measures of individual latent ability in T-maze learning. Research report 78-5 , Department of Statistics, University of Uppsala, 1978

IV. Olsson, U. On the robustness of factor analysis against crude classification of the observations. Research report 78-6 , Department of Statistics, University of Uppsala, 1978

V. Olsson, U. Maximum likelihood estimation of the polychoric correlation coefficient. Research report 78-7 , Department of Statistics, University of Uppsala, 1978

References to the papers will be made using the roman figures I - V.

SOME DATA ANALYTIC PROBLEMS IN MODELS WITH LATENT VARIABLES

by

Ulf Olsson

Doctoral dissertation to be publicly examined in room IV, the University building, University of Uppsala, on May 25, 1978, at 10 a.m., for the Degree of Doctor of Philosophy (according to Royal proclamation No 327, 1969).

Abstract

In recent years, various models for latent variables have been studied in the social sciences. In the five papers of this dissertation, different problems involving models for latent variables are studied, with an emphasis on the data analytic problems relating to these models. The first two papers deal with applications of recent statistical methods for models with latent variables to problems of psychological origin. The third paper discusses the problem of measuring learning in T-mazes from dichotomous data, using simple response models. In the last two papers, ordinal observed data are treated as indicators of latent continuous variables. In one paper, the results of factor analyses based on the observed and on the latent variables are compared. The other paper is concerned with estimating the polychoric correlation coefficient, which is the correlation between the latent variables. This is a generalization of the tetrachoric correlation coefficient to polytomous data.

Uppsala 1978

ISBN 91-506-0156-3

CONTENTS

Page	
1	Introduction
7	A longitudinal factor model for studying change in ability structure
9	Selection bias in confirmatory factor analysis
11	A comparison of measures of individual latent ability in T-maze learning
13	On the robustness of factor analysis against crude classification of the observations
15	Maximum likelihood estimation of the polychoric correlation coefficient
17	References

PREFACE

During the work with this thesis, several people have kindly shared their time, and have given me encouragement, help and support.

Professor Karl G. Jöreskog has served as my instructor for the research included in my dissertation. He has also initiated the research projects giving financial support to my work. Without doubt, his patient and continuous encouragement has been of great importance for my research.

Professor Anders Christoffersson has spent much of his time on stimulating discussions with me on scientific and other subjects. During Prof. Jöreskog´s leave of absence in the academic year 1976-77 he also served as my instructor.

My thanks are also due to my co-writer, Dr. Lars R. Bergman, now at the Swedish Central Bureau of Statistics, and to the members of the research project "Structural equation models in the social sciences", Hans Melander, B. Sc., Dr. Bengt Muthén and Dr. Dag Sörbom.

Participants in several seminars have given valuable comments on my work. In particular, I am obliged to Lena Lindén, B. Sc., Professor Einar Lyttkens, Lars Monsén, B. Sc., Docent Adam Taube, Docent Jan Vegelius and Docent Anders Ågren of the Statistics department, and to Dr. Torbjörn Järbe and Dr. Peter Thunberg of the Psychology department, all at the University of Uppsala.

The programmer of our project, Bengt Anders Dahlqvist, has supplied a valuable subroutine, used in some of my programs. The typing of the manuscripts has fast and skilfully been performed by the project´s secretary, Ami Nicholls, and by Ann-Christine Pettersson. Maggie McKay has carefully revised my English.

During the course of my work, I have received financial support from the Swedish Council for Social Science Research under project "Statistical methods for analysis of longitudinal data", and from the Bank of Sweden Tercentenary Foundation under project "Structural equation models in the

social sciences". Inspiration to some of my papers has come through my work as a statistical consultant for social scientists, first financed by the Swedish Council for Social Science Research, then by the Faculty of Social Sciences at the University of Uppsala.

Uppsala, April 1978

Ulf Olsson

INTRODUCTION *

The title of this paper contains two key-words: "data analytic" and "latent variables". Both these terms may need some further discussion, in order to clarify the aim and scope of the paper. I shall here give a brief discussion of the terms, and place the summarized papers into the appropriate context.

"Data analysis" is a term closely associated with the name of John Tukey (see e.g. Tukey 1970, 1977). Tukey uses the term as a more general concept than "statistics". One may think of data analysis as a kind of detective work, with a body of statistical data as the corpus delicti. Here, I use the term to indicate methods to explore some subject-matter area step by step, by trying alternative methods and models.

Tukey distinguishes between "exploratory" and "confirmatory" data analysis; my work may be classified into the second category. The reason for using the term at all is that I wish to emphasize the applied, rather than the theoretical side of my work. Cooley and Lohnes (1971) and Gnanadesikan (1977) have adopted the term in a similar vein.

Later in the paper, summaries of the five papers of the thesis are given. First I shall briefly discuss to what extent the papers may qualify for the notation of "data analytical".

---

* This introduction is partly based on applications for grants to the project "Structural equation models in the social sciences", forwarded to the Bank of Sweden Tercentenary Foundation. The applications were written jointly with Karl G. Jöreskog, Hans Melander, Bengt Muthén and Dag Sörbom.

Paper (I) treats an exploratory model building sequence, based on a psychological hypothesis of substantial interest. Paper (II), is likewise related to a psychological problem. Methodologically, it is concerned with the risks of making erroneous conclusions if the data are divided into groups by means of some selection variable.

One common statistical problem is to develop methods for estimating the parameters of interest. A related data analytical problem is to decide what to estimate, i.e. the choice of indicators (see Mosteller and Tukey, 1968, p 105f). In paper (III), both problems are discussed in relation to data from a certain type of learning experiments.

Another problem of concern for data analysts is the study of the robustness of different methods. One then studies what happens if the assumptions underlying the methods are not exactly met. In paper (IV), we investigate the effects on factor analysis which may follow if the observed variables are ordinal with only a few scale steps. In paper (V), finally, the problem is to estimate a correlation coefficient from ordinal data.

The second term of the title was "latent variables". This term is related to the fact that social scientists frequently work with concepts and constructs which are not directly measurable (see e.g. Torgerson, 1958, ch. 1, Goldberger, 1972 , Aigner and Goldberger, 1977).

The simplest form of a latent variable is related to the concept of reliability. Most measuring instruments suffer from lack of accuracy in the sense that independent replications of the measuring process do not always produce exactly the same results. In formal terms, a simple model for this situation may be formulated as

(1) $x = \xi + \varepsilon$.

Here, x is the observed measurement, ξ is a hypothetical (latent) true value, and ε is the measurement error, often assumed to be uncorrelated with ξ and to have zero expectation. In this sense, most variables are in reality "latent" variables.

On a more complex level, latent variables are often used to indicate concepts or constructs which cannot be reliably measured by one variable alone. Each single variable may contain not only what we want it to measure, but also other components, including measurement errors. By defining a new, latent, variable as the common core of a number of indicators, we obtain a better instrument. For example, results on different tests on addition, division and multiplication etc. may serve as indicators of a latent variable called "numerical ability", or we may say that numerical ability is a factor which determines the values of the observable variables.

This type of latent variables was first used by psychologists in factor analysis. The methodology of factor analysis is now well developed, and is described in standard textbooks (Harman, 1967; Lawley and Maxwell, 1971; Mulaik, 1972; Gorsuch, 1974).

The traditional factor analytic methods have been extended to confirmatory factor analysis (Jöreskog, 1969), and generalized to hold for several populations simultaneously (Jöreskog, 1971), and to models which permit a structure on the factor means (Sörbom, 1974). A further line of development has resulted in models and computer programs for analyzing more complicated covariance structures (Jöreskog 1970, 1973a , 1974).

Recently, there has been a strong interest in models where different latent and observable variables are connected in specified causal relations. Such models are often formulated in terms of path analysis (Wright, 1934, 1960; Duncan, 1966; Werts and Linn, 1970). Several models of this type may be analyzed within the framework of the LISREL model (Jöreskog 1973b, 1977), using the corresponding computer program (Jöreskog and Sörbom, 1976 a,1978). For applications of the methods discussed here, see e.g. Jöreskog (1976a, 1977), Sörbom (1976) and Jöreskog and Sörbom (1976b), and references therein.

Paper (I) is an application of a LISREL model, where an earlier version of the computer program (Jöreskog and van Thillo, 1973) was used. In (II), we study the effects of selection on factor analysis, within the framework provided by the methods for simultaneous factor analysis in several populations (Jöreskog, 1971), using the computer program COFAMM (Sörbom and Jöreskog, 1976).

Above, we related the concept of latent variables to simple forms of measurement errors. More complex types of measurement errors may occur because of limitations of the measuring instrument, for example such that the observed variables are qualitative. An example is when we can only record whether or not an event has occured, where occurence of the event may be coded as x=1, and non-occurence as x=0. Depending on whether or not the dichotomous variables are dependent or independent variables, this situation may be handled in different ways. One alternative is to regard the probability of the observation of x=1 as a function of some latent variable ξ, i.e.

$$P(x=1)=f(\xi). \tag{2}$$

A second alternative is to introduce a latent variable ξ which is related to x by expressions of the type

$$\begin{aligned} &x=0 \text{ if } \xi<\tau \\ &x=1 \text{ if } \xi>\tau \ , \end{aligned} \tag{3}$$

where τ is called a threshold. It may be noted that the two alternatives may be shown to be formally equivalent. See e.g. Lord and Novick (1968), Bock and Lieberman (1970), Nerlove and Press (1973), Christoffersson (1975, 1977) and Muthén (1977) for applications of these approaches to different problems. In paper (III), we use the first way of writing to study different measures of learning ability in T-maze learning.

The second approach may be generalized to hold for multinomial responses where the observed variables are ordinal. Examples of such variables are attitude items, rating scales, Likert items and the like. Data of this kind are often obtained in the behavioural sciences. A response

model for such data may be formulated as

$$
\begin{aligned}
&x=1 \text{ if } \xi < a_1 \\
&x=2 \text{ if } a_1 \leq \xi < a_2 \\
&x=3 \text{ if } a_2 \leq \xi < a_3 \\
&\quad \cdot \quad \cdot \\
&x=s \text{ if } a_{s-1} \leq \xi
\end{aligned}
\tag{4}
$$

In paper (IV), we compare the results of factor analyses based on the observed variables, x, with those obtained from ξ. In (V), we discuss estimation of the polychoric correlation coefficient, which is the correlation coefficient between the latent variables. This is based on ordinal observed data, and assumes a response model of the type (4) to hold.

I. A LONGITUDINAL FACTOR MODEL FOR STUDYING CHANGE IN ABILITY STRUCTURE

A frequently asserted hypothesis concerning intellectual development is the age differentiation hypothesis stated by Garrett (1946), "Abstract or symbol intelligence changes in its organization as age increases from a fairly unified and general ability to a loosely organized group of abilities or factors" (p. 373). In the paper, we discuss this hypothesis, mainly from a methodological point of view. For a review on earlier research in this area, reference is made to Reinert (1970) and Anastasi (1970).

Using a set of real data, originally collected for the Örebro project (Magnusson, Dunér and Zetterblom, 1975), a model building sequence is performed. The data consists of scores on six ability tests (Härnqvist, 1962) and two achievement tests, administered at ages 10 and 13 to samples of n=375 girls and n=353 boys. The two samples were analyzed separately. The model building comprised the following steps:

1. A number of unrestricted maximum likelihood factor analyses were performed, using the program UFABY3 (Jöreskog and van Thillo, 1971). The solutions did suggest a 3-factor solution at age 10 and a 4-factor solution at age 13 for both sexes.
2. Restricted factor analyses, using the hypothetical grouping of the tests, were performed using the program SIFASP (van Thillo and Jöreskog, 1970). Acceptable fit was obtained for 4-factor solutions for both ages and sexes.

3. These separate restricted solutions were combined into longitudinal models, intended to describe the development of ability structure over time. Estimation of the parameters of the models was performed using the program LISREL (Jöreskog and van Thillo, 1973). Three such models were tested:

(i) All factors were permitted to correlate freely over time, and the unique parts were assumed to be uncorrelated. This model showed a fairly poor fit to the data, and the residuals for the between-occasions part of the model were larger than the residuals for the within-occasions part.

(ii) Using a simple reparametrization of the LISREL model, the unique parts were allowed to correlate. The improvement in fit was considerable, but the solution was fairly unstable with large standard errors for the between-occasion coefficients.

(iii) The factors at age 13 were assumed to correlate directly only with the corresponding factor at age 10. Although this is a more restricted model, the extra restrictions did not significantly detrimentally affect th fit. This model was therefore chosen as a reasonable description of the data.

The parameter estimates of the chosen model do indicate that the factors are very strongly related over time. As regards the age differentiation hypothesis, the model suggests that the knowledge and the inductive factors, which are highly correlated at age 10, become more separated at age 13. This is interpreted as supporting differentiation to some extent.

II. SELECTION BIAS IN CONFIRMATORY FACTOR ANALYSIS

"Selection" is a statistical term for the procedure of restricting an investigation to some subset of the population under study. This restriction may be intended or unintended. If the selection mechanism (i.e. the procedure by which the selection is performed) is in any way related to the variables which are studied, a general observation is that the results may be biased (see e.g. Eklund, 1960).

In the paper, we study the effects of a certain type of selection on factor analysis. Following Lawley (1943-44) our selection mechanism consists of a variable (or variable vector) $\underset{\sim}{z}$. $\underset{\sim}{z}$ is assumed to correlate with the variables under study, $\underset{\sim}{v}$. The regression of $\underset{\sim}{v}$ on $\underset{\sim}{z}$ is assumed linear and homoscedastic, i.e. the conditional variances and covariances $\mathrm{Cov}(v_i, v_j|\underset{\sim}{z})$ are independent of $\underset{\sim}{z}$. A selection on $\underset{\sim}{z}$ is defined as any process which changes its density function from $\varphi(\underset{\sim}{z})$ to $\varphi^*(\underset{\sim}{z})$. For example, the selection may consist in restricting the study to individuals with specified values of $\underset{\sim}{z}$. Using these assumptions, Lawley derived relations between the parameters of the selected and the unselected populations.

We employ Lawley's results to show that, if a factor analysis is performed on a selected population:

(1) The factor matrix and the unique variances remain invariant under selection.

(2) The factor correlations are changed by the selection.

(3) The selection affects the factor means.

(4) The residuals are uncorrelated with the factors even after selection.

The first two results were also obtained by Ahmavaara (1954) and Meredith (1964). The third result is fairly selfevident. The fourth result, which was not explicitly stated by the earlier writers, implies that a factor analysis model holds even in the selected population.

The results obtained so far are used in the paper in a discussion on the so called performance differentiation hypothesis (see e.g. Reinert, 1970). This hypothesis states that individuals with a high over-all intellectual performance tend to be more differentiated with regard to their abilities. This means that a high-performing group would exhibit lower factor correlations than a low-performing group. Using the same data as in (I), it is demonstrated that differences in factor correlations observed in the data may be an artifact caused by the selection. On the other hand, if the differentiation hypothesis is true (which means that the relations between the factors and the selection variable is non-linear and/or heteroscedastic), this effect may be confounded with selection effects.

III. A COMPARISON BETWEEN MEASURES OF INDIVIDUAL LATENT ABILTY IN T-MAZE LEARNING.

In this paper, we discuss the statistical analysis of data from so-called T-maze learning experiments. In such experiments, the individual (for example an animal) is to pass through a T-shaped maze. According to specified rules, it is rewarded if it takes one way, but not if it takes the other one. Each subject makes a number of trials, and one basic idea is that the amount of training received increases the subject´s "understanding" of the rules, which in turn increases the probability of a correct response.

The vague concept of "understanding" is in the paper formalized as two learning models, (Estes 1950, 1961) where the single parameter of the models is interpreted as a latent variable describing the individual´s ability to learn. Estimation of the parameter by unweigthed least squares, generalized least squares and maximum likelihood is discussed. These estimates are compared, on theoretical and empirical grounds, with three conceptually simple measures of learning ability, defined respectively as the weighted and unweigthed sum of the correct responses, and as the waiting time to achieve k correct responses among the last n responses. The comparisons are performed with regard to the measure´s ability to differentiate between individuals with different true learning abilty, using the correlation ratio as a criterion.

In a Monte-Carlo study, all the measures are computed from a set of simulated data. The results suggest that the measures based on unweighted least squares estimation, and the waiting time to achieve k correct responses among the last n,

are susceptible to random variation. The other measures seem to perform in a comparable way, with a slight advantage for the weighted sum, at least when the number of trials is large. A set of real data is also analyzed.

IV. ON THE ROBUSTNESS OF FACTOR ANALYSIS AGAINST CRUDE CLASSIFICATION OF THE OBSERVATIONS.

Ordinal variables with only a few scale steps are rather common, especially in the behavioural sciences. When analyzing this kind of data, a common approach is to assign integer values to the categories of the variables, and proceed in the analysis as if the data had been continuous with desired distributional properties.

The paper is concerned with the consequences this approach may have on factor analysis. It is hypothesized that the observed variables are obtained through a classification of some true variables, which are multivariate normal, and for which a factor analysis model holds. Using simple formulae for the relations between true correlations and correlations based on the classified variables, the relationship between true factor models and results obtained from classified data is demonstrated. This is done numerically, using artificial population data generated with several choices of true thresholds, true factor loadings and numbers of variables, assuming a one-factor model.

The results indicate that classification may lead to a substantial lack of fit of the model, i.e. an erroneous indication that more factors are needed. This is especially true if the variables are skewed in opposite directions and have high true loadings, but does not depend much on the number of scale steps. The classification also attenuates the loading estimates, and this effect is increased with decreasing

number of scale steps and increasing variation in skewness among the variables. The number of variables does not seem to have any major impact on the results.

Extensive data scanning is recommended as a means to detect skewed variables before a factor analysis is performed.

V. MAXIMUM LIKELIHOOD ESTIMATION OF THE POLYCHORIC CORRELATION COEFFICIENT.

A common type of data, especially in the behavioural sciences, is observations on scales like

approve strongly | approve | don't know | disapprove | disapprove strongly .

Data of this kind are often quantified by assigning integer values to each category, in the example from 1 to 5. The analyses are then performed as if the data had been parametric. It may be shown that this approach may lead to biased estimates of, for example, correlation coefficients (see e.g. IV).

In the paper, we discuss estimation of the so-called polychoric correlation coefficient. It is assumed that underlying the discrete observed variables x and y are some continuous latent variables ξ and η, which are bivariate normal. These are classified into a number of classes. The observed variables are assumed to be generated so that an observation, say x=k, is obtained if the individual has a value of the latent variable ξ in the interval $a_{k-1} \leq \xi < a_k$, where a_{k-1} and a_k are class-limits. These limits are usually called thresholds. This approach is a generalization of the arguments behind the familiar tetrachoric correlation coefficient to more than two classes.

Since the thresholds are unknown parameters, the problem consists in estimating the correlation coefficient $\rho_{\xi\eta}$ and the thresholds, from a cross-tabulation of x by y. This may be done in at least two ways:

1) Simultaneous estimation of ρ and the thresholds.
2) The thresholds are first estimated as the inverse of the normal distribution function, evaluated at the cumulative marginal proportions of the table, and the estimation of ρ is then performed given the thresholds.

The first approach has been used earlier by Tallis (1962) for 2 x 2 and 3 x 3 tables. A modification of the second approach was used by Lancaster and Hamdan (1964) and Martinson and Hamdan (1975).

For both cases we derive maximum likelihood estimates of the correlation coefficient, based on a multinomial distribution of the observations. The covariance matrix of the parameter estimates is also derived for both cases. The results are used in a numerical computer study, based on true data, on simulated data and on a set of real data. It is concluded that the bias is small and that the empirical variances of the estimates are close to the theoretically derived ones for the data analyzed so far. For practical purposes, the differences between the two methods discussed seem to be small, especially when the true correlation is small.

REFERENCES

Ahmavaara, Y. The mathematical theory of factorial invariance under selection. Psychometrika, 1954, 19, 27-38.

Aigner, D. J. and Goldberger, A. S. (Eds.) Latent variables in socio-economic models. Amsterdam: North-Holland, 1977.

Anastasi, A. On the formation of psychological traits. American Psychologist, 1970, 25, 899-910.

Bock, R. D. and Lieberman, M. Fitting a response model for n dichotomously scored items. Psychometrika, 1970, 35, 179-197

Christoffersson, A. Factor analysis of dichotomized variables. Psychometrika, 1975, 40, 5-32.

Christoffersson, A. Two-step weighted least squares factor analysis of dichotomized variables. Psychometrika, 1977, 42, 433-438.

Cooley, W. W. and Lohnes, P. R. Multivariate data analysis. New York, Wiley, 1971.

Duncan, O. D. Path analysis: sociological examples. American Journal of Sociology, 1966, 72, 1-16.

Eklund. G. Studies in selection bias in applied statistics. Uppsala, Almqvist & Wiksell, 1960.

Estes, W. K. Growth and function of mathematical models for learning. In: R. Glaser (Ed.) Current trends in psychological theory. Pittsburgh, University of Pittsburgh press, 1961. Also in E. D. Neimark and W. K. Estes (Eds) Stimulus sampling theory. San Fransisco, Holden-Day, 1967.

Estes, W. K. Toward a statistical theory of learning. Psychological Review, 1950, 57, 94-104.

Garrett, H. E. A developmental theory of intelligence. American Psychologist, 1946, 1, 372-378.

Gnanadesikan, R. Methods for statistical data analysis of multivariate observations. New York, Wiley, 1977.

Goldberger, A. S. Structural equation methods in the social sciences. Econometrika, 1972, 40, 979-1001.

Gorsuch, R. L. Factor analysis. Philadelphia, Saunders, 1974.

Harman, H. H. Modern factor analysis. Chicago: The University of Chicago press, 1967.

Härnqvist, K. Manual till DBA (Manual for DBA). Stockholm, Skandinaviska Testförlaget, 1962.

Jöreskog, K. G. A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 1969, 34, 183-202.

Jöreskog, K. G. A general method for analysis of covariance structures. Biometrika, 1970, 57, 239-251.

Jöreskog, K. G. Simultaneous factor analysis in several populations. Psychometrika, 1971, 36, 409-426.

Jöreskog, K. G. Analysis of covariance structures. In: P. R. Krishnaiah (Ed.) Multivariate analysis III. New York, Academic Press, 1973, 263-285. (a)

Jöreskog, K. G. A general method for estimating a linear structural equation system. In: A. S. Goldberger and O. D. Duncan (Eds.) Structural equation models in the social sciences. New York, Seminar press, 1973, 85-112. (b)

Jöreskog, K. G. Analyzing psychological data by structural analysis of covariance matrices. In: R. C. Atkinson, D. H. Krantz, R. D. Luce and P. Suppes (Eds.) Contemporary developments in mathematical psychology, Vol. II. San Fransisco, Freeman, 1974, 1-56.

Jöreskog, K. G. Causal models in the social sciences: the need for methodological research. In: Uppsala University 500 years, 7. Faculty of Social Sciences at Uppsala University. Uppsala: Acta Universitatis Upsaliensis, 1976.

Jöreskog, K. G. Structural equation models in the social sciences: specification, estimation and testing. In: P.R. Krishnaiah (Ed.), Applications of statistics. Amsterdam, North-Holland, 1977.

Jöreskog, K. G. and van Thillo, M. New rapid algorithms for factor analysis by unweighted least squares, generalized least squares and maximum likelihood. Research memorandum 71-5. Princeton, N. J. Educational Testing Service, 1971.

Jöreskog, K. G. and van Thillo, M. LISREL - a general computer program for estimating a linear structural equations system involving multiple indicators of unmeasured variables. Research report 73-5, Department of Statistics, University of Uppsala, 1973.

Jöreskog, K. G. and Sörbom, D. LISREL-III. Estimation of linear structural equation systems by maximum likelihood methods. A Fortran IV program. Chicago, National Educational Resources, 1976 (a).

Jöreskog, K. G. and Sörbom, D. Some models and estimation methods for analysis of longitudinal data. In: D. J. Aigner and A. S. Goldberger (Eds.) Latent variables in socio-economic models. Amsterdam, North-Holland, 1976 (b).

Jöreskog, K. G. and Sörbom, D. LISREL IV: A general computer program for estimation of linear structural equation systems by maximum likelihood methods. Department of Statistics, University of Uppsala, 1978.

Lancaster, H. O. and Hamdan, M. A. Estimation of the correlation coefficient in contingency tables with possibly nonmetrical characters. Psychometrika, 1964,29,383-391.

Lawley, D. N. A note on Karl Pearsons selection formulae. Proc. Roy. Soc. Edinburgh, Section A. (Math. and Physics Section). 1943-44, 62, Part 1, 28-30.

Lawley, D. N. and Maxwell, A. E. Factor analysis as a statistical method. London: Butterworths, 1971.

Lord, F. M. and Novick, M. R. Statistical theories of mental test scores. Reading, Mass. Addison-Wesley, 1968.

Magnusson, D., Dunér, A. and Zetterblom, G. Adjustment - a longitudinal study. Stockholm, Almqvist & Wiksell, 1975.

Martinson, E. O. and Hamdan, M. A. Algorithm AS 87: Calculation of the polychoric estimate of correlation in contingency tables. Applied statistics,1975,24,272-278.

Meredith, W. Notes on factorial invariance. Psychometrika, 1964, 29, 177-185.

Mosteller, F. and Tukey, J. W. Data analysis, including statistics. In: G. Lindzey and E. Aronson (Eds.) The handbook of social psychology, second ed., vol 2. Reading, Mass., Addison-Wesley, 1968.

Mulaik, S. A. The foundations of factor analysis. New York, Mc Graw-Hill, 1972.

Muthén, B. Statistical methodology for structural equation models involving latent variables with dichotomous indicators. Doctoral dissertation, Department of Statistics, University of Uppsala, 1977.

Nerlove, M. and Press, S. J. Univariate and multivariate log-linear and logistic models. Santa Monica, The Rand Corporation, R: 1306 - EDA/NIH, 1973.

Reinert, G. Comparative factor analysis studies of intelligence throughout the human life span. In L. R. Goulet and P. B. Baltes (Eds.) Life-span developmental psychology. New York, Academic Press, 1970, 467-484.

Sörbom, D. A general method for studying differences in factor means and factor structure between groups. British Journal of Mathematical and Statistical Psychology, 1974, 27, 229-239.

Sörbom, D. Statistical methodology for model building with latent variables. Doctoral dissertation, Department of Statistics, University of Uppsala, 1976.

Sörbom, D. and Jöreskog, K.G. COFAMM: Confirmatory factor analysis with model modification. Chicago, National Educational Resources, 1976.

Tallis,G. M. The maximum likelihood estimation of correlation from contingency tables. Biometrics, 1962, 18, 342-353.

van Thillo, M. and Jöreskog, K. G. SIFASP - a general computer program for simultaneous factor analysis in several populations. Research Bulletin 70-62. Princeton, N. J., Educational Testing Service, 1970.

Tukey, J. W. Exploratory data analysis Limited preliminary edition, Reading, Mass. Addison-Wesley, 1970. (Mimeographed).

Tukey, J. W. Exploratory data analysis. Reading, Mass., Addison-Wesley, 1977.

Werts, C. E. and Linn, R. L. Path analysis: psychological examples. Psychological bulletin, 1970, 74, 193-212.

Wright, S. The method of path coefficients. Annals of Mathematical Statistics, 1934, 5, 161-215.

Wright, S. Path coefficients and path regressions: alternative or complementary concepts? Biometrics, 1960, 16, 189-202.

UPPSALA UNIVERSITET
REPRO-C HSC 1978